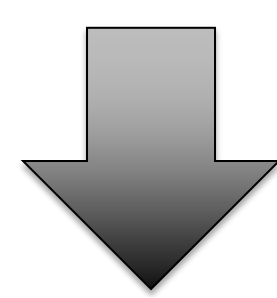
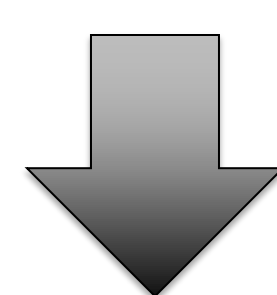


## Workflow for Evaluation and Review of DRUM Submissions

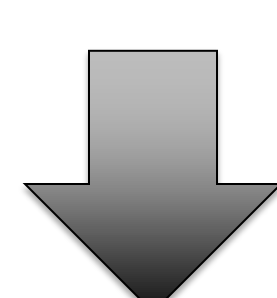
U of M Researcher submits their data to DRUM at [z.umn.edu/drum](http://z.umn.edu/drum)



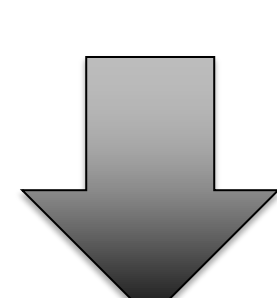
Data curators in the library receive notification that a new submission has occurred but has not yet been accepted



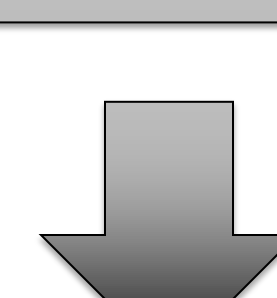
Data curators open and review files. **Identity Finder** and **Bulk Extractor** are run to uncover any personal identifying information



If restricted data is discovered, we work with the author to either remove the restricted information from the data files or reject the submission



If restricted information was successfully removed, the data is accepted in DRUM



DRUM data curators finalize the submission through additional digital preservation steps

## Introduction

The **Data Repository for the University of Minnesota (DRUM)** was launched in November 2014. DRUM's policy for depositing materials requires that data must be non-restricted and do not contain any private, confidential, or other legally protected information<sup>1,2</sup>.

To ensure that all submissions meet this criteria, several identification tools were investigated, including Identity Finder, Bulk Extractor, Find\_SSNs, and Spider. These tools were evaluated based on usability and the scope of data identified. Identity Finder and Bulk Extractor were selected for DRUM's purposes due to their extensive search and identification capabilities.

## Tools for Identifying Private Data

### Identity Finder

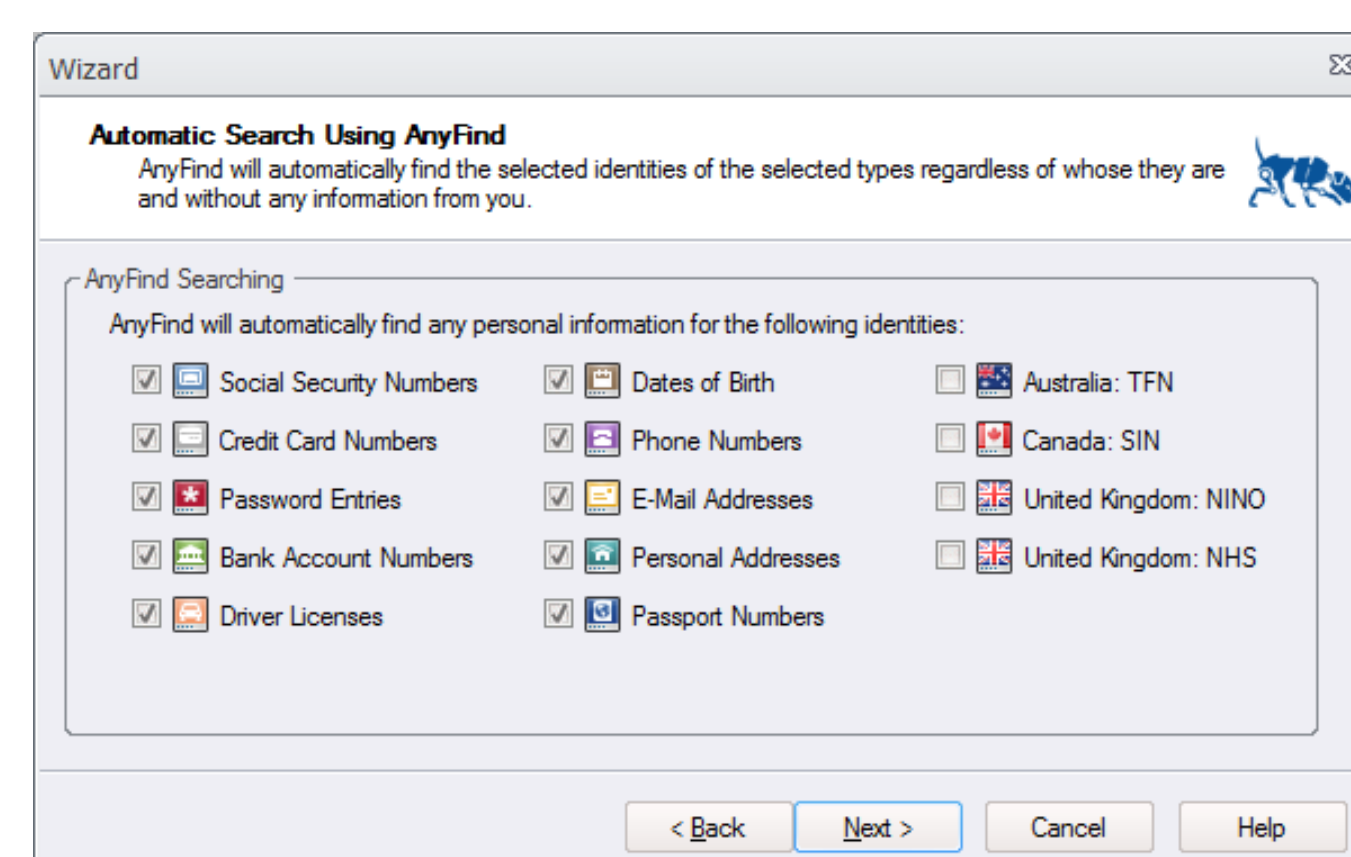
**How it works:** Searches any file type or image, stored locally or remotely on file servers, network-attached storage devices, and desktops running Windows, Mac, and Linux<sup>3</sup>.

#### Scope:

- Social security numbers
- Bank account and routing information
- Driver's licenses
- Date of birth
- Phone numbers
- Addresses
- Passports
- Credit card numbers,
- Medical record numbers
- Other national IDs

**Results:** Results are displayed through a user interface. Each identifier is listed individually with its file location and a preview window.

**Cost:** \$60 per computer



### Bulk Extractor

**How it works:** Scans a disk image, file, or directory. It can also be used to process any digital media including but not limited to hard drives, solid state drives, optical media, camera cards, cell phones, and network packet dumps<sup>6</sup>.

#### Scope:

- Credit cards
- Email addresses
- IP addresses
- Telephone numbers
- URLs
- Social security numbers
- GPS coordinates

**Results:** Creates an output directory that includes separate text files for each identifier. Also creates a histogram of the frequency and types of identifiers discovered.

**Cost:** Free

## Table of Private Identifiers

The following table lists the 19 personal identifiers as defined by the HIPAA Privacy Rule<sup>3,4</sup>. This information must be removed from any file or document before it is shared publically.

Identifier (Per UMN Policy)	Example	Review Procedure
Name	Goldy M. Gopher	Visual inspection
Geographic subdivisions smaller than a state	55 Harvard St. SW Minneapolis	Identity Finder Visual inspection
Dates directly related to individuals. All elements of dates must be removed except year.	DOB: 01/01/1980 Discharge date: 04/05/2013	Identity Finder Visual Inspection
Ages or dates indicating an age over 89	04/07/1922 Age: 96	Identity Finder Visual inspection
Telephone number	612-637-61111	Identity Finder Visual Inspection
Fax number	612-123-4567	Identity Finder Visual Inspection
Email address	gold001@umn.edu	Identity Finder Visual Inspection
Social security number	123-45-6789	Identity Finder Visual Inspection
Medical record number		Visual inspection
Health plan number	HVF123456789	Visual inspection
Account numbers	352783001	Visual inspection
Certificate/license numbers	S123-4567-8901	Identity Finder Visual inspection
Vehicle identification/serial numbers, license plate numbers	1ZVHT82H485145 PIE-314	Visual inspection
Device identification/serial numbers	W78300NC77F	Visual inspection
URLs	<a href="http://goldygphr.com">http://goldygphr.com</a>	Visual inspection Bulk Extractor
IP addresses	128.101.125.103	Visual inspection Bulk Extractor
Biometric identifiers	Finger prints	Visual inspection
Full face photographs and comparable images	MRI brain scans	Visual inspection
Any other unique identifying number, characteristics, or code	Credit card number	Visual Inspection Identity Finder

## Conclusion

The library takes several steps to prevent restricted data from being accepted into DRUM.

- Upon submission, data curators visually inspect all files and Identity Finder and Bulk Extractor are employed.
- These tools reduce the risk of exposing private data, however they come with restrictions.
- Each tool is capable of identifying only a fraction of the 19 Private Identifiers, and false negatives can also occur.

Using these tools, the workflow for the evaluation and review of submissions to DRUM was created.

## References

- [1] DRUM Policies and Terms of Use. University of Minnesota Libraries. <https://conservancy.umn.edu/pages/drum/policies/>
- [2] Research Data Management Policy: Archiving, Ownership, Retention, Security, Storage, and Transfer. UWide Policy Library. <http://policy.umn.edu/Policies/Research/RESEARCHDATA.html>
- [3] De-identifying Health Information. UWide Policy Library. [http://www.policy.umn.edu/Policies/Operations/Health/HIPAA/COMPONENT\\_PROC11.html](http://www.policy.umn.edu/Policies/Operations/Health/HIPAA/COMPONENT_PROC11.html)
- [4] The Privacy Rule. Health Information Privacy. U.S. Department of Health & Human Services. <http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/>
- [5] Identity Finder's Sensitive Data Manager. <http://www.identityfinder.com/us/>
- [6] Bulk Extractor. [http://www.forensicswiki.org/wiki/Bulk\\_extractor](http://www.forensicswiki.org/wiki/Bulk_extractor)